\|PRTS

10/501700
DT04 Rec'd PCT/PTO 16 JUL 2004

Description

Operating method for an automated language recognizer intended for the speaker-independent language recognition of words in different languages and automated language recognizer.

The method relates to an operating method of an automatic language recognizer for speaker-independent language recognition of words of different languages in accordance with Claim 1 and a corresponding automatic language recognizer in accordance with Claim 6.

For phoneme-based language recognition, a language-recognition vocabulary is necessary that contains the phonetic descriptions of all the words to be recognized. This is a basic requirement for phoneme-based language recognition. Words in this case are represented by sequences or chains of phonemes in the vocabulary. During a language recognition process, a search for the best path through the phoneme sequences in the vocabulary is carried out. This search can, for example, take place by means of the Viterbi algorithms. For continuous language recognition, the probabilities for transitions between words can also be modeled and included in the Viterbi algorithm.

The phonetic transcription for the words to be recognized form the basis of the phoneme-based language recognition. Therefore, at the start of use of a phoneme-based language recognition, the question is always how such phonetic transcripts can be obtained. Phonetic transcripts in this case means the phonetic descriptions of words from a target vocabulary. This question is particularly relevant for words that are not known to the language recognizer.

Mobile or cordless telephones are known that enable speaker-dependent name selection. A user of such a telephone must in

this case train the entries contained in the electronic telephone book of the telephone, in order to be able to subsequently use the name selection by language. Normally, no other user can use this feature because the speaker-dependent name selection is suitable for only one person, i.e. for the person who has trained the language selection. To overcome this problem, the entries in the electronic telephone book can be changed to phonetic transcripts.

To determine the phonetic transcript from a written word, for example from a telephone book entry, various approaches are known. For example, the dictating systems that are generally used with a PC should be mentioned. With dictating systems of this kind a lexicon of typically more than 10,000 words with an allocation of letter sequences to the phoneme sequences is normally stored. Because a lexicon of this kind requires a very high storage capacity, it is not practical for mobile terminal devices such as mobile or cordless telephones.

Systems are also known whereby the conversion of a word to its phonetic transcript is rule-based or takes place using specially trained neural networks. As with the lexicon, this method also has the disadvantage that the language in which the phoneme sequences to be realized must be specified. In any case, names from different languages may be present particularly in electronic telephone books. Conversion would then be impossible, or only limited, with the method described above.

For this purpose, multilingual systems for determining phoneme sequences and language recognition have been developed. These systems enable phoneme sequences to be created from different languages.

Finally there is one other solution, i.e. a user speaks the words into a language recognition system that, from these, automatically generates sequences of phonemes. For large vocabularies, and also even for just a few dozen words such as for example in an electronic telephone book with 80 entries, this is no longer acceptable for the user.

The object of this invention is therefore to propose an operating method of an automatic language recognizer for speaker-independent language recognition of words from various languages and also a corresponding automatic language recognizer that is simple to implement, is particularly suitable for use in mobile terminal devices and can be realized at reasonable cost.

The object is achieved by an operating method with the features of Claim 1 and by an automatic language recognizer with the features of Claim 6.

The invention is essentially based on the idea of determining phonetic transcripts of words for N various languages in each case and then reprocessing these and applying them to a phoneme-based monolingual language recognizer. This procedure is essentially based on the knowledge that a user of the voice recognizer normally speaks in his mother tongue. He also pronounces foreign-language words, such as names, with a mother-tongue nuance, i.e. an accent, that can be roughly modeled by a mother-tongue language recognizer. The operating method is therefore based on a language defined as the mother tongue.

Each language can thus be described with different phonemes suitable for the particular language. It is known, however, that many phonemes in different languages resemble one another. An example of this is the "p" in English and German.

This fact is utilized in multilingual language recognition. In this case a single Hidden Markov model is created for the collection of languages, by means of which several languages can be recognized simultaneously. However, this leads to a very large Hidden Markov model with a lower recognition rate than a monolingual Hidden Markov model. Furthermore, if the collection of languages is extended, for example by a further language, a new Hidden Markov model has to be created, which is very expensive. The invention avoids this necessity.

According to the invention, in a first step of the input phase for creation of a language recognition vocabulary of an operating procedure of an automated language recognizer for speaker-independent language recognition of words from various languages, particularly for the recognition of names from various languages, the phonetic transcripts of words for N various languages are determined in each case, in order to obtain N first phoneme sequences per word corresponding to N first pronunciation variants. In a second step, the similarities between the languages are utilized. To do this, a depiction of the phonemes of each language is implemented on the particular phoneme set of the mother tongue. Furthermore, in a third step the implemented depiction on the N first phoneme sequences determined in the first step is used for each word. In this way, N second phoneme sequences corresponding to N second pronunciation variants are obtained for each word. By means of the mother-tongue language recognizer, a number of N various languages can then, after creating a language-recognition vocabulary using the N second phoneme sequences per word obtained in the preceding step, be recognized for the mother-tongue language recognizer.

The invention has the following main advantages. Whereas a look-up method in a lexicon fails with mobile terminal devices because of the large memory requirement and for multilingual language recognition the set of languages was optimized, new

Hidden Markov models have to be created and optimized for each new language, by means of the grapheme/phoneme conversion into several languages in accordance with the invention, a multilingual system is created that can be implemented with relatively simple means, that is therefore particularly suitable for use in mobile terminal devices and not least can be realized at reasonable cost. For the invention, all that is essentially required in addition to the grapheme-to-phoneme conversion is a mapping, i.e. a depiction between the individual languages, as explained above. The phoneme sequence determination and the succeeding mapping or depiction normally run offline on a device, for example a mobile telephone, a personal digital assistant or personal computer with corresponding software, and are therefore time uncritical. The resources required for this can be held in a slow external memory.

Because the language recognition vocabulary created by means of the aforementioned procedure includes an N pronunciation variant for each word, the search effort during language recognition is great. To reduce this, a further step can be introduced into the process, that is performed before the creation of the language recognition vocabulary and after generation of the N second phoneme sequences per word. In this step, the N second phoneme sequences are processed corresponding to the N second pronunciation variants of each word, in that each second phoneme sequence is analyzed and classified by means of suitable distances, particularly the Levenshtein distance, and the N second phoneme sequences of each word are reduced to a few, preferably two to three phoneme sequences, particularly in that the pronunciation variants that are least similar to the pronunciation variants of the mother tongue are omitted. Simply expressed, the least important pronunciation variants are omitted by this reduction, thus reducing the search effort during language recognition.

A further reduction in cost can be achieved in that a language identification and reduction is carried out before the first

step. As part of this language identification, the probability of belonging to each of the N various languages is determined for each word to be recognized. Using the results of this language identification, the number of languages to be processed in the first step of the method is reduced, preferably to two or three different languages. This language reduction advantageously takes place in that the languages with the least probability are not further processed. For a specific word, the result of the language identification can, for example, be as follows: "German 55%, UK English 16%, US English 14%, Swedish 3%, ...". This result enables a reduction to three different languages to be made, in that Swedish is omitted, i.e. not further processed.

The determination of the phonetic transcripts in the first step of the method takes place preferably by means of at least one neural network. Neural networks have proved suitable for determining phonetic transcripts from written words, because they produce good results with regards to accuracy, and particularly with regard to the speed of processing and can be easily implemented, particularly in software.

A Hidden Markov model, particularly one that has been created for the language defined as a mother tongue, is particularly suitable for use as a mother tongue language recognizer.

The invention also relates to a language recognizer for speaker-independent language recognition of words from various languages, particularly for recognizing names from various languages. In this case, one of the various languages is defined as the mother tongue. The language recognizer includes

- a mother tongue language recognizer,

- a first processing model for determining the phonetic transcripts of words, particularly for N various languages, in order to obtain N first phoneme sequences corresponding to N first pronunciation variants per word,
- a second processing model for implementing a mapping of the phoneme of each language on the particular phoneme set of the mother tongue,
- a third processing model for applying the mapping, implemented by the second processing module, to N first phoneme sequences for each word, determined with the first processing model, whereby N second phoneme sequences corresponding to N second pronunciation variants are obtained per word, that can be recognized by the mother tongue language recognizer and
- a fourth processing model for creating a language recognition vocabulary with the N second phoneme sequences per word obtained by the third processing module for the mother tongue language recognizer.

In a preferred form of the embodiment, the automatic language recognizer has a fifth processing module for processing the N second phoneme sequences corresponding to the N second pronunciation variant of each word. The fifth processing module is designed in such a way that each second phoneme sequence is analyzed and classified using suitable distances, particularly the Levenshtein distance and the N second phoneme sequences of each word are reduced to a few, preferably two to three, phoneme sequences.

Furthermore, the automatic language recognizer can have a language identifier and a language reducer. The language identifier is connected before the first processing module and, for each word to be recognized, it determines the probability of it belonging to each of the N different languages. The language reducer reduces the number of languages to be processed by the first processing module, preferably down to two to three different languages, in that

the languages with the least probability are not further processed. The language identifier and language reducer substantially reduce both the processing effort of the automatic language recognizer, both in the input phase and in the recognition phase.

Preferably, the first processing module has at least one neural network for determining the phonetic transcripts.

Finally, the mother tongue language recognizer has, in a preferred form of embodiment, a Hidden Markov model that has been created for the language defined as the mother tongue.

Advantages and suitabilities of the invention are given in the following description of an example of an embodiment of the invention, using a single illustration. This shows a schematic flow diagram of the input phase for creation of a language recognition vocabulary in accordance with the invention.

A speaker-related name is to be selected on a mobile telephone using the names from the telephone book, for a German-speaking user. In the telephone book, there are in addition to the mainly German-language names, also some foreign-language names. A transcriber for the graphemic representation of the names is set for the German, Italian, Czech, Greek and Turkish languages, overall as N = 5 different languages.

In an initial step S0, a language identification of the supplied words 10 or entries in the telephone book is undertaken. More precisely, each individual word is analyzed with regard to the probability of it belonging to one of the five languages. If, for example, a German name is being processed, the probability for German is very high, for the

other four languages, i.e. Italian, Czech, Greek and Turkish, the probability is very much lower. Using the probabilities determined per word, the language with the lowest probability is omitted during the further processing. This means that in the succeeding processing operation there are then only four, instead of five, languages that have to be processed.

In a first step of the method S1, the phonetic transcript for each word is determined for each of the four different languages. In this way, four phoneme sequences corresponding to the four first pronunciation variants are obtained for each word.

In a second step of the method S2, a mapping of the phonemes of each of the four languages is implemented to the particular phoneme set of the mother tongue.

In a third step of the method S3, this mapping is applied to the four first phoneme sequences 12 obtained in the first step of the method S1. In this way, four second phoneme sequences 14 corresponding to the four second pronunciation variants are obtained for each word. The four second phoneme sequences 14 can already be recognized in a mother tongue language recognizer.

Furthermore to further reduce the processing effort for the language recognizer, each second phoneme sequence is analyzed and classified for each word using the Levenshtein distance (step S4). A fifth step of the method S5 then takes place, in which the analyzed and classified second phoneme sequences per word are reduced to three phoneme sequences.

Finally, in a last step S6, a language recognition vocabulary is created for the mother tongue language recognizer with the three second phoneme sequences per word obtained in the fifth step of the method S5. By again reducing the phoneme sequences

in the fifth step of the method S5, the language recognition vocabulary to be saved and to be analyzed during a language recognition process is substantially reduced. In a practical application of the language recognizer, this has the advantage on the one hand of a lower storage capacity requirement and on the other hand of a faster processing, because the vocabulary to be searched through is smaller.

After the described procedure has been completed, the user can, by means of language recognition, make a name selection, i.e. make a language-controlled call up of stored telephone numbers using the name of the subscriber, without having to once explicitly pronounce the name of the subscriber to be called, i.e. without having to "train".

The following is a brief explanation of what the user of the mobile telephone can do to improve language recognition. If he finds that a certain name is not well recognized, he can call up the language recognition menu of his mobile telephone and then select the "name selection" application. By means of this application, he can now be offered one, or several, ways of improving the language recognition of a certain word, or more precisely of a certain name, from the electronic telephone book of the mobile telephone. Some of these possibilities are briefly explained in the following by way of example.

1. The user can again speak the poorly recognized or unrecognized word into the mobile telephone and then have it converted into a phoneme sequence by means of the language recognizer contained in the mobile telephone. In this case, pronunciation variants previously automatically determined are either completely or partially, depending on their closeness to the newly determined phoneme sequence, removed from the vocabulary of the language recognizer.

2. Alternatively, the user can have a kind of phonetic transcription of the poorly recognized or unrecognized entry in the electronic telephone book shown on the display of the mobile telephone. If it is inappropriate, i.e. if there is a poor match to his pronunciation, the user can edit the kind of phonetic transcription. For example, by an automatic transcription of the entry "Jacques Chirac", "Jakwes Shirak" can be stored as a phonetic transcription. If this phonetic transcription now appears incorrect to the user, he can edit it using his mobile telephone, for example to "Zhak Shirak". The system can then also determine the phonetic description and re-enter this in the language recognition vocabulary. This should enable the automatic language recognition to function reliably.

3. Finally, the user can, by an explicit specification of a language from which a faulty or even unrecognized name originates substantially improve the recognition by an explicit selection of a specific language for a specific name. In such a case, all the pronunciation variants of the name, that are not assigned to the explicitly specified language, are removed from the language recognition vocabulary.

The invention can also be advantageously used, i.e. installed, in other mobile devices apart form a mobile telephone, e.g. a personal assistant or a personal computer.